

**UNIVERSITI TEKNOLOGI MARA**

**ANALYZING VISITOR TRENDS TO  
OPTIMIZE DATA-DRIVEN  
STRATEGIES IN PUSAT SAINS &  
KREATIVITI TERENGGANU**

**NUR SARAH BINTI ZAINOL ABIDIN**

**BACHELOR OF INFORMATION SYSTEMS  
(Hons.) BUSINESS COMPUTING**

**JULY 2025**

## **ACKNOWLEDGEMENT**

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks goes to my supervisor, Dr Hidayah for her invaluable guidance, support and encouragement throughout the course of this research. Her insightful feedback and unwavering patience have been instrumental in the successful completion of this study.

Special appreciation also goes to my beloved parents, Ibu and Abah whose unconditional love, prayers, and constant encouragement have been my greatest source of strength and inspiration. Their unwavering belief in my abilities has motivated me to strive for excellence in every endeavor.

Last but not least, I would like to give my gratitude to Sharifah Masyitah and all of my friends for their endless support, encouragement, and companionship throughout this journey. Their constant motivation and willingness to lend a helping hand whenever needed have truly made this experience more manageable and enjoyable. I am grateful to all those who contributed directly or indirectly to this research, offering their time, knowledge, and resources to ensure its completion.

Thank you for all your support and encouragement.

## ABSTRACT

This project aims to implement predictive analytics to assist Pusat Sains & Kreativiti Terengganu (PSKT) in addressing three key challenges which are limited visibility into visitor trends and demographics, underutilization of data in decision-making, and the inability to identify peak visitor periods. To overcome these issues, historical visitor data from January 2022 to December 2024 was collected and analyzed using the CRISP-DM methodology, which guided the project through business understanding, data preparation, modelling, and deployment. Three predictive algorithms including Decision Tree (DT), Random Forest (RF), and Naïve Bayes (NB) were tested to classify visitor levels into low, medium, and high categories. Among these, the Random Forest algorithm with a 70:30 data split achieved the highest accuracy of 78.48%, making it the most suitable model for integration. The predictive results were visualized through a Power BI dashboard consisting of five main pages, Home, Total Visitor Overview, Visitor Demographic, Visitor Prediction, and Visitor Impact. These visualizations enable PSKT management to explore visitor trends, understanding visitor demographic and assess the influence of external factors such as events and rentals. Expert evaluation confirmed the dashboard's clarity, usefulness, and potential as a decision-support tool. However, limitations were identified, including imbalanced data, wide value ranges, and the dashboard's lack of real-time data integration. Recommendations for future enhancement include expanding the dataset to over five years and enabling live data updates to improve prediction accuracy and extend dashboard lifespan. Overall, this project contributes a practical and scalable solution not only for PSKT but also for other institutions seeking to leverage predictive analytics for improved visitor or customer trend analysis.

# TABLE OF CONTENTS

<b>CONTENT</b>	<b>PAGE</b>
<b>SUPERVISOR APPROVAL</b>	<b>i</b>
<b>STUDENT DECLARATION</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>TABLE OF CONTENTS</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xii</b>
<b>CHAPTER ONE: INTRODUCTION</b>	<b>1</b>
1.1 Background of Study	1
1.2 Problem Statement	6
1.3 Objective	7
1.4 Scope	8
1.5 Significance	9
1.6 Project Framework	10
1.7 Gantt Chart	11
1.8 Conclusion	12
<b>CHAPTER TWO: LITERATURE VIEW</b>	<b>13</b>
2.1 Edutainment Industry	13
2.2 Business Analytics	14
2.2.1 Type of Business Analytics	15
2.2.2 Challenges in Business Analytics	16
2.2.3 Data Mining	16
2.2.4 Data Mining Techniques	17
2.3 Predictive Analytics	18
2.3.1 Predictive Analytics Algorithm	20
2.3.2 Performance Evaluation Metric	22
2.3.2.1 Accuracy	23
2.3.2.2 Confusion Matrix	23

2.4	Business Analytics Methodology	24
	2.4.1 CRISP-DM	24
	2.4.2 KDD	27
2.5	Similar Existing Research	28
2.6	Implications	30
2.7	Conclusion	31
<b>CHAPTER THREE: METHODOLOGY</b>		<b>32</b>
3.1	Project Framework	32
3.2	CRISP-DM	35
3.3	Project Planning	35
3.4	Business Understanding	36
3.5	Data Understanding	38
	3.5.1 Collect Data	38
	3.5.2 Describe Data	42
	3.5.3 Explore Data	43
	3.5.4 Verify Data	44
3.6	Data Preparation	44
	3.6.1 Select Data	45
	3.6.2 Clean Data	45
	3.6.3 Construct Data	50
	3.6.4 Integrate Data	51
	3.6.5 Format Data	53
3.7	Modelling	54
	3.7.1 Build Model	55
3.8	Evaluation	58
3.9	Deployment	59
	3.9.1 Review Project	60
	3.9.2 Expert Evaluation	60
	3.9.3 Documentation	62
3.10	Conclusion	63
<b>CHAPTER FOUR: RESULTS AND FINDINGS</b>		<b>64</b>
4.1	Current Situation of Pusat Sains & Kreativiti Terengganu (PSKT)	64
4.2	Conceptual Framework	65